Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 119 (2009) 453–467
stochastic processes and their applications
www.elsevier.com/locate/spa

An asymptotic theory for sample covariances of Bernoulli shifts

Wei Biao Wu*

Department of Statistics, The University of Chicago, 5734 S. University Avenue, Chicago, IL 60637, USA

Received 28 September 2007; received in revised form 12 February 2008; accepted 26 February 2008
Available online 4 March 2008

Abstract

Covariances play a fundamental role in the theory of stationary processes and they can naturally be estimated by sample covariances. There is a well-developed asymptotic theory for sample covariances of linear processes. For nonlinear processes, however, many important problems on their asymptotic behaviors are still unanswered. The paper presents a systematic asymptotic theory for sample covariances of nonlinear time series. Our results are applied to the test of correlations.

MSC: primary 60F05; 62M10; secondary 60G10

Keywords: Asymptotic normality; Covariance; Dependence; Linear process; Martingale; Moderate deviation; Nonlinear time series; Stationary process; Test of correlation

1. Introduction

Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary process with $\mathbb{E}(X_i^2) < \infty$; let $\mu = \mathbb{E}(X_i)$ be the mean and $\gamma_i = \mathbb{E}[(X_0 - \mu)(X_i - \mu)]$, $i \in \mathbb{Z}$, be the covariance function. Covariances characterize second order properties of the process $(X_i)_{i\in\mathbb{Z}}$ and they play a fundamental role in the theory of time series. They are critical quantities that are needed in both spectral and time domain analysis. Estimation of μ and γ_k helps understand the first and second order characteristics of the process. A typical estimate for μ is $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ and there is a rich asymptotic theory for $\sqrt{n}(\bar{X}_n - \mu)$; see [8].

* Tel.: +1 7737020958; fax: +1 7737029810.
E-mail address: wbwu@galton.uchicago.edu.

doi:10.1016/j.spa.2008.02.008

A natural estimate for γ_i is the sample covariance. For an integer $0 \le k < n$ let

$$\hat{\gamma}_k = \frac{1}{n} \sum_{i=k+1}^{n} (X_i - \bar{X}_n)(X_{i-k} - \bar{X}_n)$$

and $\hat{\gamma}_{-k} = \hat{\gamma}_k$. There exists some variants. For example, the factor $1/n$ may be replaced by $1/(n - |k|)$. Arguably, the asymptotic problem of $\hat{\gamma}_k$ is of fundamental importance in time series analysis. The latter problem has been discussed in many classical time series textbooks; see for example [6,15,1]. For other contributions see [14,17,20,23,33]. However, many of those results require that the underlying processes are linear. A challenging problem is to extend those results to nonlinear processes. In his *Reflections* [28], John Tukey commented:

> My feeling – not necessarily correct, but . . . – is that our current frequency/time techniques are quite well developed (at least so far as the present cycle goes), so that the most difficult questions are not "how to solve it" but rather either "how to formulate it", or "how do we extend applicability to less comfortable conditions".

The goal of the current paper is to present a systematic asymptotic theory for $\hat{\gamma}_k$ for nonlinear time series that are traditionally less comfortable to work with. There are many important open problems regarding the asymptotic behavior of $\hat{\gamma}_k$. For example, is there a central limit theorem (CLT) for $\hat{\gamma}_k$ with $k = k_n \to \infty$ for nonlinear processes? Can we construct simultaneous confidence intervals for $\gamma_1, \ldots, \gamma_k$ with $k \to \infty$? What is the asymptotic distribution of $\max_{i \le k_n} |\hat{\gamma}_i - \gamma_i|$? The latter can be used to test the hypothesis of white noises $\gamma_1 = \gamma_2 = \cdots = 0$. Those problems will be discussed in the paper. We will compare the performances of our test for white noises with the classical Ljung–Box test. Some open problems are also posed. The rest of the paper is organized as follows. Our main results are stated in Section 2. Some of the proofs are given in Section 3.

2. Main results

We shall work with a very general class of stationary processes which assumes the form

$$X_n = g(\ldots, \varepsilon_{n-1}, \varepsilon_n), \tag{1}$$

where ε_i are independent and identically distributed (i.i.d) innovations or shocks that drive the system [4], $\mathcal{F}_n = (\ldots, \varepsilon_{n-1}, \varepsilon_n)$ is the input, $g(\cdot)$ is a real-valued function which can be interpreted as a filter or a transform and $X_n = g(\mathcal{F}_n)$ is the output or response. The process (X_n) is causal in the sense that X_n only depends on $\mathcal{F}_n$, the innovations up to n, and it does not depend on future innovations. As argued in [30,24,27,31], the class of processes that follow within the framework of (1) is huge. It includes linear processes and a large class of nonlinear time series; see Section 5 in [26]. For discussions on non-causal two-sided processes see Remark 3.

To work with processes of form (1), we need to introduce appropriate dependence measures. Following [31], we adopt the physical dependence measure: let $(\varepsilon'_i)_{i\in\mathbb{Z}}$ be an i.i.d copy of $(\varepsilon_i)_{i\in\mathbb{Z}}$ and let

$$\delta_p(n) = \|g(\mathcal{F}_n) - X'_n\|_p, \quad \text{where } X'_n = g(\mathcal{F}'_n), \quad \mathcal{F}'_n = (\mathcal{F}_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_n). \tag{2}$$

Here we write $Z \in \mathcal{L}^p$, $p > 0$, if $\|Z\|_p := [\mathbb{E}(|Z|^p)]^{1/p} < \infty$, and $\|Z\| = \|Z\|_2$. Another coupling scheme is to use $X^*_n = g(\mathcal{F}^*_n)$, where $\mathcal{F}^*_n = (\ldots, \varepsilon'_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_n)$; see Remark 2. Let $\kappa_p = \|X_i\|_p$. Note that $\delta_p(n)$ measures the functional dependence of X_n on ε_0 and it is directly related to the data generating mechanism of the underlying process. In many applications it is convenient to work with $\delta_p(n)$; see Remarks 1 and 2. All our results below are based on the physical dependence measure $\delta_p(n)$.

2.1. CLT with bounded lags

To state central limit theorems for $\hat{\gamma}_k$, we need the following stability condition:

$$\sum_{i=0}^{\infty} \delta_p(i) < \infty. \tag{3}$$

The above condition means that the cumulative impact of innovation ε_0 on future values $(X_i)_{i \geq 0}$ is finite, thus suggesting short-range dependence. Theorem 1 provides a central limit theorem for $\hat{\gamma}_k$ with bounded lags, while Theorem 2 concerns unbounded lags.

Theorem 1. *Let $k \in \mathbb{N}$ be fixed and $\mathbb{E}(X_i) = 0$; let $Y_i = (X_i, X_{i-1}, \ldots, X_{i-k})^{\mathrm{T}}$ and $\Gamma_k = (\gamma_0, \gamma_1, \ldots, \gamma_k)^{\mathrm{T}}$. Assume $X_i \in \mathcal{L}^4$ and (3) holds with $p = 4$. Then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i Y_i - \Gamma_k) \Rightarrow N[0, \mathbb{E}(D_0 D_0^{\mathrm{T}})] \tag{4}$$

where $D_0 = \sum_{i=0}^{\infty} \mathcal{P}_0(X_i Y_i) \in \mathcal{L}^2$ and $\mathcal{P}_0$ is the projection operator defined by

$$\mathcal{P}_i \xi = \mathbb{E}(\xi | \mathcal{F}_i) - \mathbb{E}(\xi | \mathcal{F}_{i-1}), \quad i \in \mathbb{Z}.$$

Proof of Theorem 1. We shall apply the coupling argument and [16] CLT for stationary processes. Observe that $X_i' = X_i$ if $i \leq -1$. Since $\varepsilon_0', \varepsilon_j, j \in \mathbb{Z}$, are i.i.d,

$$\mathbb{E}(X_i X_{i-k} | \mathcal{F}_{-1}) = \mathbb{E}(X_i' X_{i-k}' | \mathcal{F}_{-1}) = \mathbb{E}(X_i' X_{i-k}' | \mathcal{F}_0).$$

Since $\delta_j = 0$ if $j \leq -1$, by Jensen's and the triangle inequalities and (3),

$$\begin{aligned}
\sum_{i=0}^{\infty} \|\mathcal{P}_0(X_i X_{i-k})\| &= \sum_{i=0}^{\infty} \|\mathbb{E}(X_i X_{i-k} - X_i' X_{i-k}' | \mathcal{F}_0)\| \\
&\leq \sum_{i=0}^{\infty} \|X_i X_{i-k} - X_i' X_{i-k}'\| \\
&\leq \sum_{i=0}^{\infty} (\|X_i - X_i'\|_4 \|X_{i-k}\|_4 + \|X_i'\|_4 \|X_{i-k} - X_{i-k}'\|_4) \\
&\leq \sum_{i=0}^{\infty} [\delta_4(i) + \delta_4(i-k)] \kappa_4 = 2\kappa_4 \sum_{i=0}^{\infty} \delta_4(i) < \infty. \qquad (5)
\end{aligned}$$

Hence $\sum_{i=0}^{\infty} \|\mathcal{P}_0(X_i Y_i)\| < \infty$ and, by the Cramer–Wold device, (4) follows from Theorem 1 in [16]. $\diamond$

As an immediate consequence of Theorem 1, we have the following corollary which gives a central limit theorem for sample correlations. Let $\hat{\rho}_k = \hat{\gamma}_k / \hat{\gamma}_0$, $\rho_k = \gamma_k / \gamma_0$ and $S_n =$

$X_1 + \cdots + X_n$. Assume $\mu = 0$. By Theorem 1 in [32], for $p \geq 2$,

$$\|S_n\|_p \leq C_p \sqrt{n} \sum_{i=0}^{\infty} \delta_p(i) = O(\sqrt{n}), \tag{6}$$

where C_p is a constant only depending on p. By the triangle inequality,

$$\left\| \sum_{i=k+1}^{n} X_i X_{i-k} - n\hat{\gamma}_k \right\|_p \leq 2\|\bar{X}_n\|_p \|S_{n-k}\|_p + (n-k)\|\bar{X}_n\|_p^2 = O(1). \tag{7}$$

We omit the proof of Corollary 1 since it easily follows from the Cramer–Wold device, the Slutsky theorem, and (7).

Corollary 1. *Under conditions of Theorem 1, we have for some nonnegative definite matrix W that*

$$\sqrt{n}[(\hat{\rho}_1, \ldots, \hat{\rho}_k)^{\mathrm{T}} - (\rho_1, \ldots, \rho_k)^{\mathrm{T}}] \Rightarrow N(0, W). \tag{8}$$

Remark 1. For a linear process $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$ with $\varepsilon_t \in \mathcal{L}^4$, we have $\delta_4(i) = |a_i| \|\varepsilon_0 - \varepsilon_0'\|_4$ and (3) reduces to $\sum_{i=0}^{\infty} |a_i| < \infty$, a classical condition for linear processes to be short-range dependent. In this case, for fixed $k \in \mathbb{N}$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i X_{i-k} - \gamma_k) \Rightarrow N(0, \sigma_k^2),$$

where σ_k^2 can be computed from the Bartlett formula; see [2] and Section 7.2 in [6]. However, the argument in the latter book heavily depends on the linearity assumption and it can not be easily generalized to nonlinear time series. For example, the argument does not seem to work well for the nonlinearly transformed process $G(X_t) = |X_t| - \mathbb{E}|X_t|$. In comparison, for the latter process our Theorem 1 is also applicable and (3) similarly reduces to $\sum_{i=0}^{\infty} |a_i| < \infty$. $\diamond$

Remark 2. Shao and Wu [26] provided examples of nonlinear time series that satisfy the geometric-moment contraction (GMC) property: let $(\varepsilon_i')_{i \in \mathbb{Z}}$ be an i.i.d copy of $(\varepsilon_i)_{i \in \mathbb{Z}}$; let $X_n^* = g(\mathcal{F}_n^*)$, where $\mathcal{F}_n^* = (\ldots, \varepsilon_{-1}', \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_n)$. The GMC property says that there exists $\rho \in (0, 1)$ such that

$$\|X_n - X_n^*\|_p = O(\rho^n). \tag{9}$$

It is easily seen that (9) implies $\delta_p(n) = O(\rho^n)$. So Theorem 1 is applicable. A simple example of nonlinear time series for which (9) holds is the autoregressive process

$$X_n = R(X_{n-1}, \varepsilon_n), \tag{10}$$

where ε_n are i.i.d and $R(\cdot, \cdot)$ is a bivariate measurable function such that, for some x_0,

$$R(x_0, \varepsilon_k) \in \mathcal{L}^p \quad \text{and} \quad \max_{x \neq x'} \frac{\|R(x, \varepsilon_k) - R(x', \varepsilon_k)\|_p}{|x - x'|} < 1. \tag{11}$$

Using the technique in [34], we conclude that (11) implies (9). Many nonlinear processes are of form (10); see [34,31]. $\diamond$

Remark 3. A careful check of the proof of Theorem 1 is that it also holds for two-sided non-causal processes. Specifically, let the two-sided process

$$X_n = g(\ldots, \varepsilon_{n-1}, \varepsilon_n, \varepsilon_{n+1}, \ldots) =: g(\mathcal{F}_n), \quad \text{where } \mathcal{F}_n = (\ldots, \varepsilon_{n-1}, \varepsilon_n, \varepsilon_{n+1}, \ldots).$$

As in (2), we can similarly define $\mathcal{F}'_n = (\mathcal{F}_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_n, \varepsilon_{n+1}, \ldots)$ by replacing ε_0 in $\mathcal{F}_n$ by ε'_0. Let $\delta_p(n) = \|g(\mathcal{F}_n) - g(\mathcal{F}'_n)\|_p$. Under the condition $\sum_{i\in\mathbb{Z}} \delta_4(i) < \infty$, (4) still holds with $D_0 = \sum_{i\in\mathbb{Z}} \mathcal{P}_0(X_i Y_i) \in \mathcal{L}^2$. The key step in the proof is that, instead of using Hannan's [16] CLT for causal processes, one can apply Volný's [29] CLT for non-causal processes. The details are omitted since it does not involve essential extra difficulties. Theorem 2 in Section 2.2 is also valid for two-sided processes. I would like to thank a referee for clarifying this. Since many time series encountered in practice are causal, we decide to state our theorems in the setting of causal processes. $\diamond$

Remark 4. A careful check of the proof of Theorem 1 suggests that it can be easily generalized to higher order moments and the same argument is applicable. Let $g_j(u_1, \ldots, u_k)$ be polynomials with degree $d \in \mathbb{N}$, $1 \le j \le J$, and $g = (g_1, \ldots, g_J)^{\mathrm{T}}$. Let

$$\bar{g}_n = \frac{1}{n}\sum_{i=0}^{n-1} g(X_{i+1}, \ldots, X_{i+k}). \tag{12}$$

Assume that (3) holds with $p = 2d$. Following the proof of Theorem 1, we have by the delta method that the CLT $\sqrt{n}[\bar{g}_n - \mathbb{E}(\bar{g}_n)] \Rightarrow N(0, \Sigma_g)$ holds, where Σ_g is a covariance matrix. As a special case, the latter result can be applied in the estimation of joint cumulants. For example, let $\mu = 0$ and consider the estimation of the fourth order joint cumulant

$$c_0 = \mathbb{E}(X_1X_2X_3X_4) - \mathbb{E}(X_1X_2)\mathbb{E}(X_3X_4) - \mathbb{E}(X_1X_3)\mathbb{E}(X_2X_4) - \mathbb{E}(X_1X_4)\mathbb{E}(X_2X_3),$$

we can let $g_1(u_1, \ldots, u_4) = u_1u_2u_3u_4$, $g_2(u_1, \ldots, u_4) = u_1u_2$, $g_3(u_1, \ldots, u_4) = u_1u_3$ and $g_4(u_1, \ldots, u_4) = u_1u_4$. Consequently c_0 can be estimated by the plug-in rule in view of (12). The CLT for the estimate follows from Slutsky's Theorem. $\diamond$

Remark 5. In Theorem 1, (3) is a short-range dependence condition. If (3) fails, then the process (X_i) is long-range dependent and the limiting distribution of $\hat{\gamma}_k - \gamma_k$, with possibly non-$\sqrt{n}$ normalization, may no longer be Gaussian; see [25]. $\diamond$

2.2. CLT with unbounded lags

Theorem 1 states a CLT for $\sqrt{n}(\hat{\gamma}_k - \gamma_k)$ with bounded k. It turns out that, for unbounded k, the asymptotic behavior is quite different. By Theorem 3.1 in [21], one can have a CLT for strong mixing processes with $k_n = o(\log n)$. An open problem was posed in the latter paper that whether the severe restriction $k_n = o(\log n)$ can be relaxed. The latter restriction excludes many important applications. Harris, McCabe and Leybourne [18] considered linear processes with larger ranges of k_n. Theorem 2(ii) gives a CLT for short-range dependent nonlinear processes under a natural and mild condition on k_n: $k_n \to \infty$ and $k_n/n \to 0$.

Theorem 2. *Let $Z_i = (X_i, X_{i-1}, \ldots, X_{i-h+1})^{\mathrm{T}}$, where $h \in \mathbb{N}$ is fixed. Let $k_n \to \infty$, $\mathbb{E}(X_i) = 0$ and assume* (3) *with $p = 4$. Then we have* (i)

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[X_i Z_{i-k_n} - \mathbb{E}(X_{k_n} Z_0)] \Rightarrow N(0, \Sigma_h), \tag{13}$$

where Σ_h is an $h \times h$ matrix with entries

$$\sigma_{ab} = \sum_{j \in \mathbb{Z}} \gamma_{j+a}\gamma_{j+b} = \sum_{j \in \mathbb{Z}} \gamma_j \gamma_{j+b-a} =: \sigma_{0,a-b}, \quad 1 \le a, b \le h, \tag{14}$$

and (ii) *if additionally $k_n/n \to 0$, then*

$$\sqrt{n}[(\hat{\gamma}_{k_n}, \ldots, \hat{\gamma}_{k_n-h+1})^{\mathrm{T}} - (\gamma_{k_n}, \ldots, \gamma_{k_n-h+1})^{\mathrm{T}}] \Rightarrow N(0, \Sigma_h). \tag{15}$$

An interesting observation of Theorem 2 is that the asymptotic covariance matrix Σ_h in (13) does not depend on the speed of $k_n \to \infty$. As an immediate corollary, we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i X_{i-k_n} - \gamma_{k_n}) \Rightarrow N(0, \sigma_{00})$$

and $\sigma_{00} = \sum_{j \in \mathbb{Z}} \gamma_j^2$ does not depend on k_n. The latter property is quite different from the one in which k is bounded. It turns out that, as expected, $\lim_{k \to \infty} \| \sum_{i=0}^{\infty} \mathcal{P}_0(X_i X_{i-k}) \|^2 = \sigma_{00}$; see Remark 6 in Section 3.

The covariance matrix Σ_h in (14) has an interesting structure. Let $\eta_i, i \in \mathbb{Z}$, be i.i.d standard Gaussian random variables and

$$G_i = \sum_{j \in \mathbb{Z}} \gamma_j \eta_{i-j}. \tag{16}$$

Then the covariance matrix of the Gaussian vector $(G_1, \ldots, G_h)^{\mathrm{T}}$ is Σ_h. Theorem 2 describes an interesting fact that, asymptotically, $n^{-1/2} \sum_{i=1}^{n} [X_i Z_{i-k_n} - \mathbb{E}(X_{k_n} Z_0)]$ behaves like $(G_1, \ldots, G_h)^{\mathrm{T}}$. For the special case of linear processes, a similar claim was made in Section 7.2 of [6].

2.3. Supremum of sample covariances

Assume $k_n \to \infty$ and $k_n/n \to 0$. The problem of obtaining the asymptotic distribution of $\max_{1 \le k \le k_n} |\hat{\gamma}_k - \gamma_k|$ is very difficult and, on the other hand, is extremely important since the result can be used to construct simultaneous confidence intervals of $\{\gamma_k, k \ge 1\}$ and can be used to test the hypothesis of white noises: $\gamma_1 = \gamma_2 = \ldots = 0$. Here, with [12,13] martingale moderate deviation results and a [32] martingale approximation scheme, we are able to obtain an asymptotic upper distributional bound for $\max_{1 \le k \le k_n} |\hat{\gamma}_k - \gamma_k|$. However, the asymptotic distribution of the latter is still unknown (cf Conjecture 1). Let

$$\breve{\gamma}_j = \frac{1}{n} \sum_{i=1}^{n} X_{i-j} X_i. \tag{17}$$

Let $\phi = \Phi'$ the standard normal density function and let Φ^{-1} be the inverse of Φ.

Theorem 3 asserts the speed of normal approximation in the form of moderate deviation for $\sqrt{n}(\breve{\gamma}_j - \gamma_j)$. For $x \ge 1$ let ι_x be the solution to the equation

$$x = (1 + \iota_x)^9 \exp(\iota_x^2/2). \tag{18}$$

As $x \to \infty$, one has the expansion $\iota_x^2 = 2 \log x - (18 + o(1)) \log(1 + \sqrt{2 \log x})$; see [12]. The proof of Theorem 3 is given in Section 3.2.

Theorem 3. *Assume that $\delta_8(i) = O(i^{-\beta})$ with $\beta > 3/2$, $k = o(n)$, and that there exists a constant $c_0 > 0$ such that $\sigma_k = \|\sum_{i=0}^{\infty} \mathcal{P}_0 X_{i-k} X_i\| \ge c_0$. Let $b_n = k/n + n^{-2/3}$. Then*

$$\left| \frac{\mathbb{P}(\sqrt{n}|\breve{\gamma}_k - \gamma_k|/\sigma_k > \iota_x)}{2[1 - \Phi(\iota_x)]} - 1 \right| \le \theta_0 (x b_n)^{1/5} \tag{19}$$

uniformly over $x \in [1, 1/b_n]$, where θ_0 is a constant independent of k and n.

Corollary 2. *Assume that $k_n = o[n^{1/2}(\log n)^{-2}]$, $\min_{k \le k_n} \sigma_k > c_0$ for some $c_0 > 0$, $\delta_8(i) = O(i^{-\beta})$ with $\beta > 3/2$. Let $0 < \alpha < 1$. Then*

$$\limsup_{n\to\infty} \mathbb{P}\left[\sqrt{n} \max_{k \le k_n} \frac{|\hat{\gamma}_k - \gamma_k|}{\sigma_k} > \Phi^{-1}(1 - \alpha/(2k_n))\right] \le \alpha. \tag{20}$$

Proof of Corollary 2. Let $S_m = \sum_{i=1}^{m} X_i$, $\Delta_{n,k} = \sum_{i=1+k}^{n} X_i X_{i-k} - (n-k)\gamma_k$ and

$$R_{n,k} = n\hat{\gamma}_k - \sum_{i=1+k}^{n} X_i X_{i-k} = (n-k)\bar{X}_n^2 - \bar{X}_n S_{n-k} - \bar{X}_n (S_n - S_k) - k\gamma_k.$$

Since $\delta_8(i) = O(i^{-\beta})$, $\beta > 3/2$, $X_k = \sum_{j\in\mathbb{Z}} \mathcal{P}_j X_k$ and $\mathcal{P}_j$ are orthogonal, for $k \in \mathbb{N}$,

$$\begin{aligned} |\gamma_k| &= \left| \mathbb{E} \sum_{j,j'\in\mathbb{Z}} \mathcal{P}_j X_0 \mathcal{P}_{j'} X_k \right| = \left| \sum_{j\in\mathbb{Z}} \mathbb{E}(\mathcal{P}_j X_0 \mathcal{P}_j X_k) \right| \\ &\le \sum_{j\in\mathbb{Z}} \|\mathcal{P}_j X_0\| \|\mathcal{P}_j X_k\| \le \sum_{j=-\infty}^{0} \delta_2(-j)\delta_2(k-j) = O(k^{-\beta}). \end{aligned}$$

So $\sup_{k\ge1} |k\gamma_k| < \infty$. By (6), we have $\|R_{n,k}\| = O(1)$. Let $t_n = n^{1/4}$, $u_n = \Phi^{-1}(1 - \alpha/(2k_n))$ and $u'_n = (1 - k_n/n)^{-1/2} u_n \pm t_n (n - k_n)^{-1/2}/\sigma_k$. Elementary calculations show that, for either sign in u'_n, we have $\Phi(-u'_n)/\Phi(-u_n) \to 1$ and, for all large n, $(1 + u'_n)^9 \exp(u'^2_n/2) \le n^2/(nk_n + n^{4/3})$. Applying Theorem 3 to $\Delta_{n,k}$ and $\iota_x = u'_n$, we have

$$\begin{aligned} \mathbb{P}(\sqrt{n}|\hat{\gamma}_k - \gamma_k| > \sigma_k u_n) &= \mathbb{P}(|R_{n,k} + \Delta_{n,k}| > \sigma_k u_n n^{1/2}) \\ &\le \mathbb{P}(|\Delta_{n,k}| > \sigma_k u_n n^{1/2} - t_n) + \mathbb{P}(|R_{n,k}| > t_n) \\ &= k_n^{-1}\alpha(1 + o(1)) + O(t_n^{-2}) = k_n^{-1}\alpha(1 + o(1)). \end{aligned}$$

So (20) easily follows. ◇

Recall (14) for σ_{ab}. In Section 2.2, for large k, (16) gives an asymptotic probabilistic representation for $\sqrt{n}(\hat{\gamma}_k - \gamma_k)$ in terms of the Gaussian process G_k. We conjecture that (20) in Corollary 2 can be improved to (21) below.

Conjecture 1. *Assume that $k_n = o[n^{1/2}(\log n)^{-2}]$, $k_n \to \infty$, and $\delta_8(i) = O(i^{-\beta})$ with $\beta > 3/2$. Let $0 < \alpha < 1$. Then for every $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{P}\left[\sqrt{n} \max_{k \le k_n} |\hat{\gamma}_k - \gamma_k| \le \sigma_{00}^{1/2}(a_{k_n} x + b_{k_n})\right] = \exp(-2\exp(-x)), \tag{21}$$

where $b_k = (2\log k - \log\log k - \log(4\pi))^{1/2}$ and $a_k = 1/b_k$.

The intuition is that the appropriately normalized maximum of the stationary Gaussian process $G_i/\sigma_{00}^{1/2}$, $i = 1, \dots, m$, is the same as that of independent standard Gaussian random variables under the assumption that $\mathbb{E}(G_0 G_m) \log m \to 0$; see [19] and [3]. Here a_k and b_k come from the extreme value theory of Gaussian processes.

2.4. Inference of covariances

Corollary 2 can be used to construct *simultaneous confidence intervals* for covariances. Given $X_1, \dots, X_n$, we can construct Bonferroni-type conservative simultaneous confidence intervals for $\gamma_1, \dots, \gamma_{k_n}$ with $k_n = o[n^{1/2}(\log n)^{-2}]$ by

$$\hat{\gamma}_k \pm \hat{\sigma}_k \frac{\Phi^{-1}(1-\alpha/(2k_n))}{\sqrt{n}}, \quad k = 1, \dots, k_n, \tag{22}$$

where $\hat{\sigma}_k$ are estimates of σ_k. Proposition 1(i) below asserts $\mathcal{L}^1$ consistency of an estimate of σ_k, and Proposition 1(ii) is for $\sum_{j=-l_n}^{l_n} \hat{\gamma}_j^2$, an estimate for $\sum_{j\in\mathbb{Z}} \gamma_j^2$. Its proof is given in Section 3.3. The imposed conditions are not the weakest possible. The problem of choosing an optimal lag l_n using data-driven approaches is beyond the scope of the paper.

Proposition 1. *Let $l_n \in \mathbb{N}$ satisfy $l_n \to \infty$ and $l_n = o(\sqrt{n})$.* (i) *Let $k \in \mathbb{N}$ be fixed. Assume (3) with $p = 8$. Define $L_i = X_i X_{i-k} - \gamma_k$, $L_i^\diamond = (X_i - \bar{X}_n)(X_{i-k} - \bar{X}_n) - \hat{\gamma}_k$, $\nu_i = \mathrm{cov}(L_0, L_i)$, $\nu_i^\diamond = (n-i-k)^{-1} \sum_{l=1+k}^{n+i} L_l^\diamond L_{l+i}^\diamond$, and $\varsigma^\diamond = \nu_0^\diamond + 2\sum_{i=1}^{l_n} \nu_i^\diamond$. Then* $\lim_{n\to\infty} \mathbb{E}|\varsigma^\diamond - \sigma_k^2| = 0$. (ii) *Assume (3) with $p = 4$. Then* $\lim_{n\to\infty} \mathbb{E}|\sum_{j=-l_n}^{l_n} \hat{\gamma}_j^2 - \sum_{j\in\mathbb{Z}} \gamma_j^2| = 0$.

The simultaneous confidence intervals (22) are conservative in the sense that, by (20) of Corollary 2, the asymptotic coverage probability is no less than $1-\alpha$. The latter conservativeness issue does not seem to be overly severe. Let $\alpha = 0.05$ and $k_n = 10$. As an idealization, assume that $\sqrt{n}(\hat{\gamma}_k - \gamma_k)$ are asymptotically independent. Then 100% minus the probability in the left hand side of (20) is $\{2\Phi[\Phi^{-1}(1-\alpha/(2k_n))] - 1\}^{k_n} = (1-\alpha/k_n)^{k_n} = 0.95111$, which is quite close to the nominal level $1 - \alpha = 0.95$.

An important problem in the inference of stochastic processes is to test whether a process is a white noise sequence. For example, after a model is fitted to some observed data, one would like to inspect the residuals and perform model diagnostics. If the residuals do not behave like a white noise sequence, then one may need to find a better model which can capture more structural information from the data.

Various tests for white noise have been proposed in the literature and they include Fisher's test, generalized likelihood ratio test, χ^2-test and Neyman test; see Section 7.4 in [10]. Let X_i, $i \in \mathbb{Z}$, be a stationary sequence. The null hypothesis is

$$H_0: \quad \gamma_1 = \gamma_2 = \dots = 0. \tag{23}$$

Here we shall compare the performance of [5] portmanteau test and the one based on Corollary 2. The Ljung–Box test statistic has the form

$$Q_{LB} = n(n+2) \sum_{k=1}^{k_n} \frac{\hat{\rho}_k^2}{n-k}, \quad \text{where } \hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}. \tag{24}$$

Let $\alpha \in (0, 1)$. If $Q_{LB} > \chi^2_{k_n,1-\alpha}$, the $(1-\alpha)$th quantile of χ^2 distribution with k_n degrees of freedom, then at least one of the γ_ks is different from zero at level α. The simultaneous confidence intervals (22) suggest a natural test for H_0: if 0 is within every interval, then we accept H_0 at level α, otherwise we reject it.

To compare the powers of the two tests, consider the simple example $X_i = (1-\theta)^{1/2}\varepsilon_i + \theta\varepsilon_{i-1}$, where $\theta \in (-1, 1)$ and ε_i are i.i.d standard normal random variables. Then $\gamma_1 = (1-\theta)^{1/2}\theta$ and $\gamma_k = 0$ for $k \geq 2$. Let $k_n = \lfloor n^{1/3} \rfloor$. If $\theta = \theta_n \to 0$ satisfies $[(\log k_n)/n]^{1/2} = o(\theta_n)$, or equivalently $\theta_n^{-2} \log \theta_n^{-2} = o(n)$, then the power of our test approaches 1. In comparison, the Ljung–Box test requires the stronger condition $(k_n/n)^{1/2} = o(\theta_n)$, or $\theta_n^{-3} = o(n)$, to ensure that the power goes to 1. In other words, for a given $\theta \neq 0$ which is close to 0, the sample size n needed in the Ljung–Box test to reject the independence hypothesis is much larger than the one needed in our test. So our test has a better power. Similar conclusions hold for the AR process $X_i = \theta X_{i-1} + \varepsilon_i$.

3. Proofs

Recall (2) for $\mathcal{F}'_i$, $X'_i = g(\mathcal{F}'_i)$ and $\kappa_p = \|X_i\|_p$. This section provides proofs for Theorems 2 and 3 and Proposition 1.

3.1. Proof of Theorem 2

(i). We first prove (13) with $h = 1$. Let $\tilde{X}_i = \mathbb{E}(X_i|\mathcal{F}_{i-l,i})$, where $l \in \mathbb{N}$ and $\mathcal{F}_{m,n} = (\varepsilon_m, \varepsilon_{m+1}, \ldots, \varepsilon_n)$, and $\tau_l = \|X_i - \tilde{X}_i\|_4$. By the triangle and the Schwarz inequalities,

$$\begin{aligned}\|X_i X_{i-k_n} - \tilde{X}_i \tilde{X}_{i-k_n}\| &\leq \|X_i X_{i-k_n} - X_i \tilde{X}_{i-k_n}\| + \|X_i \tilde{X}_{i-k_n} - \tilde{X}_i \tilde{X}_{i-k_n}\| \\ &\leq \|X_i\|_4 \|X_{i-k_n} - \tilde{X}_{i-k_n}\|_4 + \|X_i - \tilde{X}_i\|_4 \|\tilde{X}_{i-k_n}\|_4 \leq 2\kappa_4 \tau_l.\end{aligned}$$

By (5), $\|\mathcal{P}_0 X_i X_{i-k_n}\| \leq [\delta_4(i) + \delta_4(i-k_n)]\kappa_4$. Clearly, $\|\mathcal{P}_0 \tilde{X}_i \tilde{X}_{i-k_n}\| = 0$ if $i < 0$. Since $k_n \to \infty$ and $\tilde{X}_i$ and $\tilde{X}_{i-k_n}$ are independent if $k_n > l$, $\|\mathcal{P}_0 \tilde{X}_i \tilde{X}_{i-k_n}\| = 0$ if $i \geq k_n > l$. Note that $\mathcal{P}_0 \tilde{X}_i = \mathbb{E}(\mathcal{P}_0 X_i|\mathcal{F}_{i-l,i})$. Then if $0 \leq i < k_n$, $\|\mathcal{P}_0 \tilde{X}_i \tilde{X}_{i-k_n}\| = \|\tilde{X}_{i-k_n} \mathcal{P}_0 \tilde{X}_i\| \leq \kappa_4 \delta_4(i)$. So, for all large n,

$$\begin{aligned}\|\mathcal{P}_0(X_i X_{i-k_n} - \tilde{X}_i \tilde{X}_{i-k_n})\| &\leq \min(\|X_i X_{i-k_n} - \tilde{X}_i \tilde{X}_{i-k_n}\|, \|\mathcal{P}_0 X_i X_{i-k_n}\| + \|\mathcal{P}_0 \tilde{X}_i \tilde{X}_{i-k_n}\|) \\ &\leq \min\{2\kappa_4 \tau_l,\ 2[\delta_4(i) + \delta_4(i-k_n)]\kappa_4\}.\end{aligned}$$

Let $Q_n = \sum_{i=1}^n X_i X_{i-k_n} - n\gamma_{k_n}$ and $\tilde{Q}_n = \sum_{i=1}^n \tilde{X}_i \tilde{X}_{i-k_n} - n\mathbb{E}(\tilde{X}_0 \tilde{X}_{k_n})$. Then

$$\begin{aligned}\limsup_{n\to\infty} \frac{\|Q_n - \tilde{Q}_n\|}{\sqrt{n}} &\leq \limsup_{n\to\infty} \sum_{i=0}^{\infty} \|\mathcal{P}_0(X_i X_{i-k_n} - \tilde{X}_i \tilde{X}_{i-k_n})\| \\ &\leq \limsup_{n\to\infty} \sum_{i=0}^{\infty} \min\{2\kappa_4 \tau_l, 2[\delta_4(i) + \delta_4(i-k_n)]\kappa_4\} \\ &\leq \sum_{i=0}^{\infty} 4\kappa_4 \min(\tau_l, \delta_4(i)) \to_{l\to\infty} 0. \end{aligned} \tag{25}$$

Here we have applied the inequality $\min(|a|, |b| + |c|) \leq \min(|a|, |b|) + \min(|a|, |c|)$, and the Lebesgue dominated convergence theorem since $\tau_l \to 0$ as $l \to \infty$. By (25), it remains to show that, for fixed $l \in \mathbb{N}$, $\tilde{Q}_n/\sqrt{n} \Rightarrow N(0, s_l^2)$ for some $s_l^2 < \infty$. Let $3l < k_n$ and

$$\tilde{D}_j = \sum_{i=j}^{j+l} \tilde{X}_{i-k_n} \mathcal{P}_j \tilde{X}_i = \sum_{q=0}^{l} \tilde{X}_{q+j-k_n} \mathcal{P}_j \tilde{X}_{q+j} \quad \text{and} \quad \tilde{M}_n = \sum_{j=1}^{n} \tilde{D}_j.$$

Then $\tilde{D}_j$, $j = 1, \ldots, n$, are martingale differences with respect to $\mathcal{F}_j$. Since $\tilde{X}_i = \sum_{j=i-l}^{i} \mathcal{P}_j \tilde{X}_i$,

$$\|\tilde{M}_n - \tilde{Q}_n\| = \left\| \sum_{j=1}^{n} \sum_{i=j}^{j+l} \tilde{X}_{i-k_n} \mathcal{P}_j \tilde{X}_i - \sum_{i=1}^{n} \sum_{j=i-l}^{i} \tilde{X}_{i-k_n} \mathcal{P}_j \tilde{X}_i \right\| = O(1). \tag{26}$$

We shall apply the martingale central limit theorem (cf [7]) for $\tilde{M}_n/\sqrt{n}$. Since $\|\mathcal{P}_j \tilde{X}_i\| \le \delta_4(i-j)$, $\|\tilde{D}_j\|_4 \le \kappa_4 \sum_{i=0}^{\infty} \delta_4(i)$, the Lindeberg condition is satisfied. We now show that $n^{-1} \sum_{i=1}^{n} \mathbb{E}(\tilde{D}_j^2 | \mathcal{F}_{i-1})$ converges. For $0 \le q, q' \le l$, let

$$A_j = \tilde{X}_{q+j-k_n} \tilde{X}_{q'+j-k_n} \quad \text{and} \quad B_j = (\mathcal{P}_j \tilde{X}_{q'+j})(\mathcal{P}_j \tilde{X}_{q+j}).$$

Since $3l < k_n$, A_j and B_j are independent, and A_j and $\mathcal{P}_{j-t} B_j$ $(1 \le t \le l)$ are also independent. For $t = 1, \ldots, l$, we have, by the orthogonality of $A_j(\mathcal{P}_{j-t} B_j)$,

$$\left\| \sum_{j=1}^{n} A_j (\mathcal{P}_{j-t} B_j) \right\| = \sqrt{n} \|A_1(\mathcal{P}_{1-t} B_1)\| = O(\sqrt{n}). \tag{27}$$

Since $\mathbb{E}(B_j | \mathcal{F}_{j-l-1}) = \mathbb{E}(B_j)$ and $\mathbb{E}(B_j | \mathcal{F}_{j-1}) - \mathbb{E}(B_j | \mathcal{F}_{j-l-1}) = \sum_{t=1}^{l} \mathcal{P}_{j-t} B_j$, by (27),

$$\begin{aligned}
&\mathbb{E} \left| \sum_{i=1}^{n} A_j \mathbb{E}(B_j | \mathcal{F}_{j-1}) - n\mathbb{E}(A_j)\mathbb{E}(B_j) \right| \\
&\quad \le \mathbb{E} \left| \sum_{i=1}^{n} A_j [\mathbb{E}(B_j | \mathcal{F}_{j-1}) - \mathbb{E}(B_j | \mathcal{F}_{j-l-1})] \right| + |\mathbb{E}(B_1 | \mathcal{F}_{-l})| \mathbb{E} \left| \sum_{i=1}^{n} A_j - n\mathbb{E}(A_j) \right| \\
&\quad = O(\sqrt{n}) + o(n)
\end{aligned}$$

by the ergodic theorem and (27). Hence

$$\frac{1}{n} \sum_{j=1}^{n} \mathbb{E}(\tilde{D}_j^2 | \mathcal{F}_{j-1}) \to s_l^2 := \sum_{q,q'=0}^{l} \tilde{\gamma}_{q-q'} \mathbb{E}[(\mathcal{P}_0 \tilde{X}_{q'})(\mathcal{P}_0 \tilde{X}_q)] \text{ in probability.}$$

By the martingale central limit theorem, $\tilde{M}_n/\sqrt{n} \Rightarrow N(0, s_l^2)$. Observe that

$$\begin{aligned}
s_l^2 &= \sum_{q,q' \in \mathbb{Z}} \tilde{\gamma}_{q-q'} \mathbb{E}[(\mathcal{P}_0 \tilde{X}_{q'})(\mathcal{P}_0 \tilde{X}_q)] = \sum_{q \in \mathbb{Z}} \sum_{i \in \mathbb{Z}} \tilde{\gamma}_i \mathbb{E}[(\mathcal{P}_0 \tilde{X}_q)(\mathcal{P}_0 \tilde{X}_{q+i})] \\
&= \sum_{i \in \mathbb{Z}} \tilde{\gamma}_i \sum_{q \in \mathbb{Z}} \mathbb{E}[(\mathcal{P}_{-q} \tilde{X}_0)(\mathcal{P}_{-q} \tilde{X}_i)] = \sum_{i \in \mathbb{Z}} \tilde{\gamma}_i^2.
\end{aligned} \tag{28}$$

With elementary manipulations, we have $s_l^2 \to \sum_{i \in \mathbb{Z}} \gamma_k^2$ as $l \to \infty$. Hence, by (25) and (26), $Q_n/\sqrt{n} \Rightarrow N(0, \sigma_{00})$.

Using the Crámer-Wold device, one can apply the above argument and obtain (13) for the general case $h \ge 2$. We omit the details.

(ii). It easily follows from (i) in view of (6) and (7), Slutsky's Theorem and the fact that $\sqrt{n}/\sqrt{n-k_n} \to 1$. $\diamond$

Remark 6. For a fixed integer $k \geq 0$, by Theorem 1, we have $\sqrt{n}(\hat{\gamma}_k - \gamma_k) \Rightarrow N(0, \sigma_k^2)$, where $\sigma_k = \| \sum_{i=0}^{\infty} \mathcal{P}_0(X_i X_{i-k}) \|$. The proof of Theorem 2 implies that

$$\lim_{k \to \infty} \sigma_k^2 = \sum_{j \in \mathbb{Z}} \gamma_j^2. \tag{29}$$

To this end, recall $\tilde{X}_i = \mathbb{E}(X_i | \mathcal{F}_{i-l,i})$. Let $\tilde{\sigma}_k = \| \sum_{i=0}^{\infty} \mathcal{P}_0 \tilde{X}_i \tilde{X}_{i-k} \|$. By (25), we have

$$\begin{aligned} |\sigma_k - \tilde{\sigma}_k| &\leq \sum_{i=0}^{\infty} \| \mathcal{P}_0(X_i X_{i-k_n} - \tilde{X}_i \tilde{X}_{i-k_n}) \| \leq \sum_{i=0}^{\infty} 2\kappa_4 \min\{\tau_l, \delta_4(i) + \delta_4(i - k_n)\} \\ &\leq \sum_{i=0}^{\infty} 4\kappa_4 \min(\tau_l, \delta_4(i)) \to_{l \to \infty} 0. \end{aligned} \tag{30}$$

For k with $k > 3l$, we have $\tilde{\sigma}_k = \| \sum_{i=0}^{l} \mathcal{P}_0 \tilde{X}_i \tilde{X}_{i-k} \| = \| \tilde{D}_0 \|$. Since $\| \tilde{D}_0 \|^2 = s_l^2 = \sum_{i \in \mathbb{Z}} \tilde{\gamma}_i^2$, by (30), we have (29). $\diamond$

3.2. *Proof of Theorem 3*

To prove Theorem 3, we need Lemma 1 which concerns moderate deviations of stationary processes. The argument in the proof of Theorem 1 in [35] can be modified to show Lemma 1, which is a consequence of Theorem 2.1 in [12] (see also [13]). Since there are no essential difficulties involved, we omit the details.

Lemma 1. *Let $D_i, i \in \mathbb{Z}$, be stationary martingale differences with respect to the filter $\mathcal{F}_i$. Assume $D_i \in \mathcal{L}^4$ and $\sigma = \| D_0 \| > 0$. Let $M_n = \sum_{i=1}^{n} D_i$, $V_n = \sum_{i=1}^{n} \mathbb{E}(D_i^2 | \mathcal{F}_{i-1})$ and*

$$I_n := \sum_{i=1}^{n} \frac{\| D_i \|_4^4}{n^2} + \| V_n/n - \sigma^2 \|_2^2 = \| D_i \|_4^4/n + \| V_n - n\sigma^2 \|_2^2/n^2. \tag{31}$$

Let $b_n > 0$ be a sequence such that $I_n = O(b_n)$ and let R_n be a random sequence such that

$$\| R_n \|_4^4/n^2 = o(b_n). \tag{32}$$

Then there exists a constant C, independent of x and n, such that

$$\left| \frac{\mathbb{P}(M_n + R_n \geq \sqrt{n}\sigma r_x)}{1 - \Phi(r_x)} - 1 \right| + \left| \frac{\mathbb{P}(M_n + R_n \leq -\sqrt{n}\sigma r_x)}{\Phi(-r_x)} - 1 \right| \leq C(x b_n)^{1/5} \tag{33}$$

holds uniformly in $x \in [1, b_n^{-1}]$.

Proof of Theorem 3. Let $(\varepsilon_i')_{i \in \mathbb{Z}}$ be an i.i.d copy of $(\varepsilon_i)_{i \in \mathbb{Z}}$. Let

$$\eta_p(n) = \| X_n - X_n^* \|_p, \text{ where } X_n^* = g(\mathcal{F}_n^*) \quad \text{and} \quad \mathcal{F}_n^* = (\ldots, \varepsilon_{-1}', \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_n). \tag{34}$$

By Theorem 1(iii) in [31],

$$\eta_8^2(n) \leq 64 \sum_{i=-\infty}^{0} \delta_8^2(n - i) = O(n^{1-2\beta}). \tag{35}$$

By Jensen's inequality, since $\mathbb{E}(X_n^*|\mathcal{F}_0) = \mathbb{E}(X_n^*) = 0$, $\|\mathbb{E}(X_n|\mathcal{F}_0)\|_8 = \|\mathbb{E}(X_n - X_n^*|\mathcal{F}_0)\|_8 = O(n^{1/2-\beta})$. Note that $1/2 - \beta < -1$. Hence

$$\sum_{i=0}^{k-1} \|X_{i-k}\mathbb{E}(X_i|\mathcal{F}_0) - \gamma_k\| \le \sum_{i=0}^{k-1} \|X_{i-k}\|_4 \|\mathbb{E}(X_i|\mathcal{F}_0)\|_4 = O(1)$$

and, since $\mathbb{E}(X_{i-k}^* X_i^*|\mathcal{F}_0) = \gamma_k$,

$$\begin{aligned}\sum_{i=k}^{\infty} \|\mathbb{E}(X_{i-k}X_i|\mathcal{F}_0) - \gamma_k\| &\le \sum_{i=k}^{\infty} \|\mathbb{E}(X_{i-k}X_i - X_{i-k}^* X_i^*|\mathcal{F}_0)\| \\ &\le \sum_{i=k}^{\infty} (\eta_8(i-k)\|X_i\|_4 + \|X_{i-k}^*\|_4 \eta_8(i)) = O(1).\end{aligned}$$

Observe that the preceding two inequalities hold uniformly over $k \ge 0$. So

$$\sup_{k\ge 0} \sum_{i=0}^{\infty} \|\mathbb{E}(X_{i-k}X_i|\mathcal{F}_0) - \gamma_k\| = O(1). \tag{36}$$

We now apply the argument in [11] to approximate $n(\breve{\gamma}_k - \gamma_k)$ by a martingale. For $m \in \mathbb{Z}$ define

$$H_m = \sum_{i=m}^{\infty} [\mathbb{E}(X_{i-k}X_i|\mathcal{F}_m) - \gamma_k] \quad \text{and} \quad D_m = H_m - \mathbb{E}(H_m|\mathcal{F}_{m-1}).$$

Then $D_1, \dots, D_n$ are martingale differences and $X_{m-k}X_m = H_m - \mathbb{E}(H_{m+1}|\mathcal{F}_m)$. Let $M_n = \sum_{i=1}^n D_i$, $V_n = \sum_{i=1}^n \mathbb{E}(D_i^2|\mathcal{F}_{i-1})$ and $R_n = n(\breve{\gamma}_k - \gamma_k) - M_n$. So

$$\|R_n\|_4 = \|n(\breve{\gamma}_k - \gamma_k) - M_n\|_4 \le 2\|H_m\|_4 = O(1). \tag{37}$$

Next we shall apply Lemma 1 to prove (19). To this end, a key step is to obtain a bound for $\|V_n - n\sigma_k^2\|$. Since $\mathbb{E}(D_i^2|\mathcal{F}_{i-1}) = \sum_{j=1}^{2k-1} \mathcal{P}_{i-j} D_i^2 + \mathbb{E}(D_i^2|\mathcal{F}_{i-2k})$, we have

$$\|V_n - n\sigma_k^2\| \le \sum_{j=1}^{2k-1} \left\| \sum_{i=1}^n \mathcal{P}_{i-j} D_i^2 \right\| + \left\| \sum_{i=1}^n \mathbb{E}(D_i^2|\mathcal{F}_{i-2k}) - n\sigma_k^2 \right\| =: K_n + J_n.$$

For K_n, since $\mathcal{P}_{i-j} D_i^2$, $i = 1, \dots, n$, are martingale differences, by Schwarz's inequality,

$$K_n = \sqrt{n} \sum_{j=1}^{2k-1} \|\mathcal{P}_{1-j} D_1^2\| \le \sqrt{2kn} \left(\sum_{j=1}^{2k-1} \|\mathcal{P}_{1-j} D_1^2\|^2 \right)^{1/2} \le \sqrt{2kn} \|D_1^2\|.$$

The treatment for the second term J_n is more complicated. For $i \ge 0$ let

$$\lambda_i := \|X_i X_{i-k} - X_i' X_{i-k}'\|_4 \le \|X_i\|_8 \|X_{i-k} - X_{i-k}'\|_8 + \|X_{i-k}'\|_8 \|X_i - X_i'\|_8 \tag{38}$$

and

$$\Lambda_i := \|X_i X_{i-k} - X_i^* X_{i-k}^*\|_4 \le \|X_i\|_8 \|X_{i-k} - X_{i-k}^*\|_8 + \|X_{i-k}^*\|_8 \|X_i - X_i^*\|_8. \tag{39}$$

By Proposition 3 in [32], for $1 \le i \le n$,

$$\|\mathbb{E}(D_{i+2k}^2|\mathcal{F}_0) - \sigma_k^2\| \le 8\|D_0\|_4 \Lambda_{i+2k} + 8\|D_0\|_4 \sum_{j=i+2k}^{\infty} \min(2\Lambda_{j+1}, 2\lambda_{j+1-(i+2k)})$$
$$\le 8\|D_0\|_4 \Lambda_{i+2k} + 16\|D_0\|_4 \sum_{j=1}^{\infty} \min(\Lambda_{j+i+2k}, \lambda_j).$$

Choose $d \in \mathbb{N}$ such that $2^{d-1} < n \le 2^d$. By Proposition 2.3 in [22], there exists an absolute constant c such that

$$J_n = \left\| \sum_{i=1}^{n} \mathbb{E}(D_{i+2k}^2|\mathcal{F}_0) - n\sigma_k^2 \right\|$$
$$\le c\sqrt{n}\|D_0^2\| + c\sqrt{n} \sum_{i=1}^{n} i^{-1/2} \|\mathbb{E}(D_{i+2k}^2|\mathcal{F}_0) - \sigma_k^2\|$$
$$= O(\sqrt{n}) + O(\sqrt{n}) \sum_{i=1}^{n} i^{-1/2} \left[\Lambda_{i+2k} + \sum_{j=1}^{\infty} \min(\Lambda_{i+2k}, \lambda_j) \right].$$

Note that $\lambda_j \le \|X_0\|_8 \delta_8(j)$ if $1 \le j < k$, and $\lambda_j = O((j-k+1)^{-\beta})$ if $j \ge k$. Elementary calculations show that, as $\epsilon \downarrow 0$, $\sum_{j=1}^{k-1} \min(\epsilon, \lambda_j) = O(\epsilon^{1-1/\beta})$, and also $\sum_{j=k}^{\infty} \min(\epsilon, \lambda_j) = O(\epsilon^{1-1/\beta})$. By (35) and (39), since $\beta > 3/2$, we have

$$J_n = O(\sqrt{n}) + O(\sqrt{n}) \sum_{i=1}^{n} i^{-1/2} O(\Lambda_{i+2k}^{1-1/\beta})$$
$$= O(\sqrt{n}) + O(\sqrt{n}) \sum_{i=1}^{n} i^{-1/2} (i+k)^{(1/2-\beta)(1-1/\beta)} = O(n^{2/3}).$$

Hence $\|V_n - n\sigma_k^2\| = O((kn)^{1/2} + n^{2/3})$. We now apply Lemma 1 with $b_n = k/n + n^{-2/3}$. For I_n in (31), we have $I_n = O(b_n)$. By (37), (32) trivially holds. So (33) implies (19). $\diamond$

3.3. *Proof of Proposition 1*

We first prove (ii). Assume without loss of generality that $\mu = 0$. For $j \ge 0$ let $\hat{\gamma}_j^\circ = n^{-1} \sum_{i=j+1}^{n} X_i X_{i-j}$ and $\hat{\gamma}_{-j}^\circ = \hat{\gamma}_j^\circ$. Then by (5),

$$n\|\hat{\gamma}_j^\circ - \mathbb{E}\hat{\gamma}_j^\circ\|/\sqrt{n-j} \le \sum_{i=0}^{\infty} \|\mathcal{P}_0(X_i X_{i-j})\| = O(1).$$

By (7), $n\|\hat{\gamma}_j^\circ - \hat{\gamma}_j\| = O(1)$. Since $|\mathbb{E}\hat{\gamma}_j^\circ - \gamma_j| = |j\gamma_j/n| = O(l_n/n)$, we have $\|\hat{\gamma}_j - \gamma_j\| = O(n^{-1/2})$ and hence $\mathbb{E}|\hat{\gamma}_j^2 - \gamma_j^2| = O(n^{-1/2})$. So (ii) follows since $\sum_{j \ge |l_n|} \gamma_j^2 \to 0$.

Now we prove (i). Let $\hat{\nu}_i = (n-i-k)^{-1} \sum_{l=1+k}^{n+i} L_l L_{l+i}$. Similarly as (5), we have

$$\sum_{l \in \mathbb{Z}} \|\mathcal{P}_0(X_l X_{l-k} X_{l+i} X_{l+i-k})\| \le 4\kappa_8^3 \sum_{l=0}^{\infty} \delta_8(l) < \infty.$$

Since (3) holds with $p = 8$. Hence by (5) we have uniformly in $i = 0, \ldots, l_n$ that

$$\sqrt{n-i-k}\|\hat{\nu}_i - \nu_i\| \le \sum_{l=-i}^{\infty} \|\mathcal{P}_0 L_l L_{l+i}\| = O(1). \tag{40}$$

By (6), $\|\bar{X}_n\|_4 = O(n^{-1/2})$. So $\|L_i - L_i^\diamond\| \le 2\|\bar{X}_n\|_4\|X_i\|_4 + \|\bar{X}_n^2\| + \|\hat{\gamma}_k - \gamma_k\| = O(n^{-1/2})$ and hence $\mathbb{E}|\nu_i^\diamond - \hat{\nu}_i| = O(n^{-1/2})$. By (40), $\mathbb{E}|\varsigma^\diamond - (\nu_0 + 2\sum_{i=1}^{l_n} \nu_i)| = O(l_n n^{-1/2}) = o(1)$. By (5) and Corollary 1 in [9], $\sigma_k^2 = \sum_{i\in\mathbb{Z}} \nu_i$. So (i) follows since $l_n \to \infty$ and $\sum_{i\in\mathbb{Z}} |\nu_i| < \infty$. To see the latter, since $L_i = \sum_{h\in\mathbb{Z}} \mathcal{P}_h L_i$, by the orthogonality of $\mathcal{P}_h$,

$$\nu_i = \mathbb{E}(L_0 L_i) = \mathbb{E} \sum_{h,h'\in\mathbb{Z}} (\mathcal{P}_h L_0)(\mathcal{P}_{h'} L_i) = \sum_{h\in\mathbb{Z}} \mathbb{E}(\mathcal{P}_h L_0)(\mathcal{P}_h L_i)$$

and hence by (5) $\sum_{i\in\mathbb{Z}} |\nu_i| \le \sum_{h\in\mathbb{Z}} \|\mathcal{P}_h L_0\| \sum_{i\in\mathbb{Z}} \|\mathcal{P}_h L_i\| < \infty$. $\diamond$

Acknowledgements

The authors are grateful to the referees for many helpful comments.

References

[1] T.W. Anderson, The Statistical Analysis of Time Series, Wiley, New York, 1971.
[2] T.W. Anderson, A.M. Walker, On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process, Ann. Math. Statist. 35 (1964) 1296–1303.
[3] S. Berman, Limit theorems for the maximum term in stationary sequences, Ann. Math. Statist. 35 (1964) 502–516.
[4] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, Time Series Analysis: Forecasting and Control, Prentice-Hall, NJ, 1994.
[5] G.E.P. Box, G.M. Ljung, On a measure of a lack of fit in time series models, Biometrika 65 (1978) 297–303.
[6] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, 2nd ed., Springer, New York, 1991.
[7] Y.S. Chow, H. Teicher, Probability Theory, 2nd ed., Springer, New York, 1988.
[8] J. Dedecker, P. Doukhan, G. Lang, J.R. Leon R., S. Louhichi, C. Prieur, Weak Dependence: With Examples and Applications, Springer, New York, 2007.
[9] J. Dedecker, F. Merlevède, D. Volný, On the weak invariance principle for non-adapted sequences under projective criteria, J. Theoret. Probab. 20 (2007) 971–1004.
[10] J. Fan, Q. Yao, Nonlinear Time Series: Nonparametric and Parametric Methods, Springer, New York, 2003.
[11] M.I. Gordin, The central limit theorem for stationary processes, Dokl. Akad. Nauk SSSR 188 (1969) 739–741.
[12] I.G. Grama, On moderate deviations for martingales, Ann. Probab. 25 (1997) 152–183.
[13] I.G. Grama, E. Haeusler, An asymptotic expansion for probabilities of moderate deviations for multivariate martingales, J. Theoret. Probab. 19 (2006) 1–44.
[14] P. Hall, C.C. Heyde, Martingale Limit Theorem and its Application, Academic Press, New York, 1980.
[15] E.J. Hannan, Multiple Time Series, Wiley, New York, 1970.
[16] E.J. Hannan, Central limit theorems for time series regression, Z. Wahrsch. Verw. Gebiete 26 (1973) 157–170.
[17] E.J. Hannan, The asymptotic distribution of serial covariances, Ann. Statist. 4 (1976) 396–399.
[18] D. Harris, B. McCabe, S. Leybourne, Some limit theory for autocovariances whose order depends on sample size, Econometric Theory 19 (2003) 829–864.
[19] H.C. Ho, T. Hsing, On the asymptotic joint distribution of the sum and maximum of stationary normal random variables, J. Appl. Probab. 33 (1996) 138–145.
[20] J.R.M. Hosking, Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series, J. Econom. 73 (1996) 261–284.
[21] D.M. Keenan, A central limit theorem for $m(n)$ autocovariances, J. Time Ser. Anal. 18 (1997) 61–78.
[22] M. Peligrad, S. Utev, A new maximal inequality and invariance principle for stationary sequences, Ann. Probab. 33 (2005) 798–815.
[23] P.C.B. Phillips, V. Solo, Asymptotics for linear processes, Ann. Statist. 20 (1992) 971–1001.
[24] M.B. Priestley, Nonlinear and Nonstationary Time Series Analysis, Academic Press, London, 1988.

[25] M. Rosenblatt, Independence and dependence, in: Proc. 4th Berkeley Sympos. Math. Statist. and Prob., vol II, Univ. California Press, Berkeley, CA, 1961, pp. 431–443.
[26] X. Shao, W.B. Wu, Asymptotic spectral theory for nonlinear time series, Ann. Statist. 35 (2007) 1773–1801.
[27] H. Tong, Non-linear Time Series: A Dynamic System Approach, Oxford University Press, Oxford, 1990.
[28] Tukey, J. Reflections, et al., in: D. Brillinger (Ed.), New directions in time series analysis, Part I, in: IMA, Volumn 45, Springer, NY, 1992, pp. 387–389.
[29] D. Volný, Approximating martingales and the central limit theorem for strictly stationary processes, Stochastic Process. Appl. 44 (1993) 41–74.
[30] N. Wiener, Nonlinear Problems in Random Theory, MIT Press, Cambridge, MA, 1958.
[31] W.B. Wu, Nonlinear system theory: Another look at dependence, Proc. Nat. Journal Acad. Sci. USA 102 (2005) 14150–14154.
[32] W.B. Wu, Strong invariance principles for dependent random variables, Ann. Probab. 35 (2007) 2294–2320.
[33] W.B. Wu, W. Min, On linear processes with dependent innovations, Stochastic Process. Appl. 115 (2005) 939–958.
[34] W.B. Wu, X. Shao, Limit Theorems for Iterated Random Functions, J. Appl. Probab. 41 (2004) 425–436.
[35] W.B. Wu, Z. Zhao, Moderate deviations for stationary processes, Statist. Sinica (2007) (in press).